MEDS

Malware Evolution Discovery System

Agenda

- **Personal Background**
- **Mobile Malware**
  - **Geographical Mobile Markets**
- Intro to **Android** and **APK** package
- **MEDS** (Malware Evolution Discovery System)
  - Creation Phylogenetic (**Lineage**) Trees
  - Predicting **Generativeness**
- **Summary/Future Work**

whoami

- **Antonio Cesar Vargas**
  - **M.Sc**, John Jay College of Criminal Justice
    - Digital Forensics and Cybersecurity
  - **B.Sc**, Queens College
    - Computer Science
  - Interests
    - **Python**
    - **Malware**
    - **Memory Forensics**

2014 10th Birthday Mobile Malware

Mobile Malware

- **Android**
  - Prefered Malware Creators
  - Dominant Mobile Platform
- **Threats**
  - Ransomware, botnet, personal/financial information theft

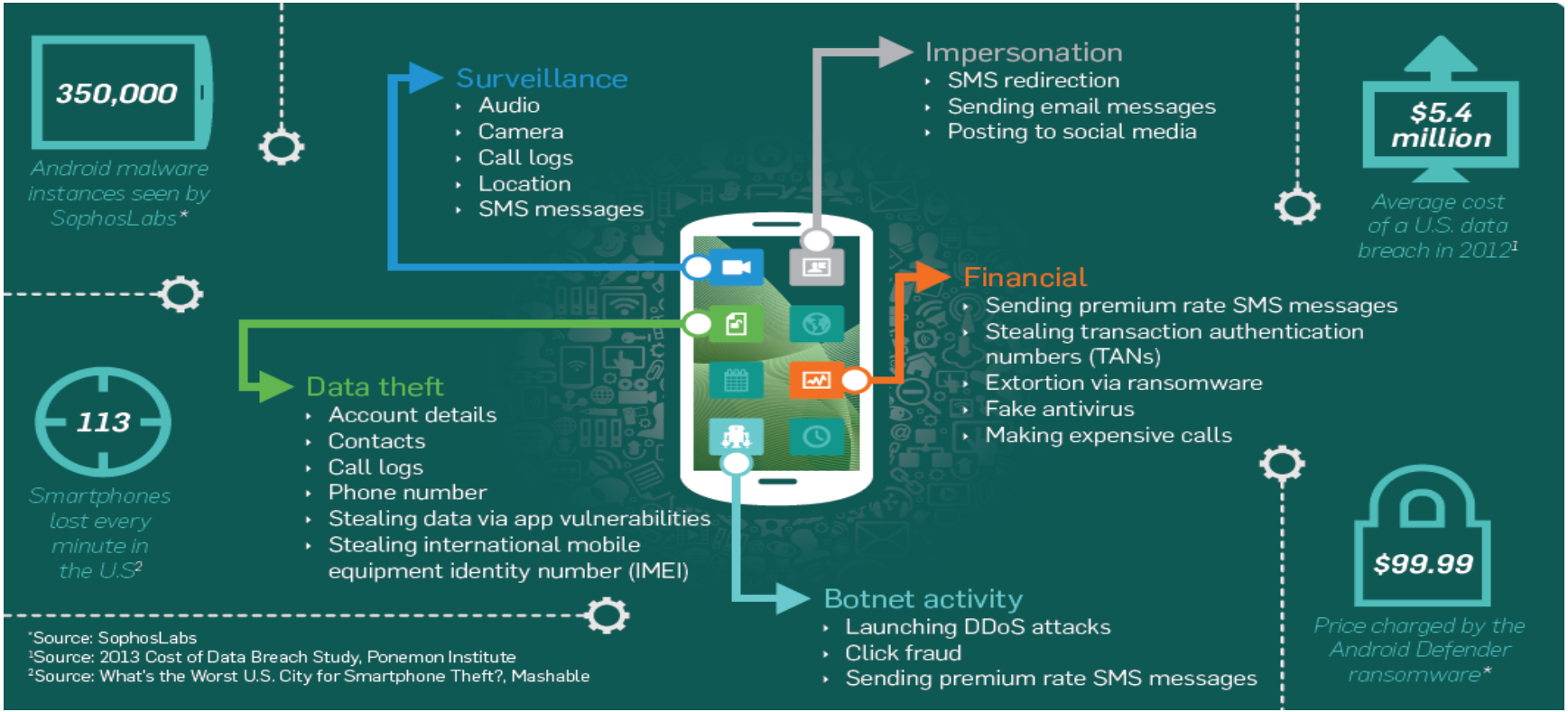
Tenth Anniversary Mobile Malware

- **2004**
  - Symbian OS
    - Cabir, Trojan.Mos, Skulls
- **2006**
  - Cross Platform Mobile Malware(Symbian and Blackberry)
    - Redbrowser, FlexiSpy

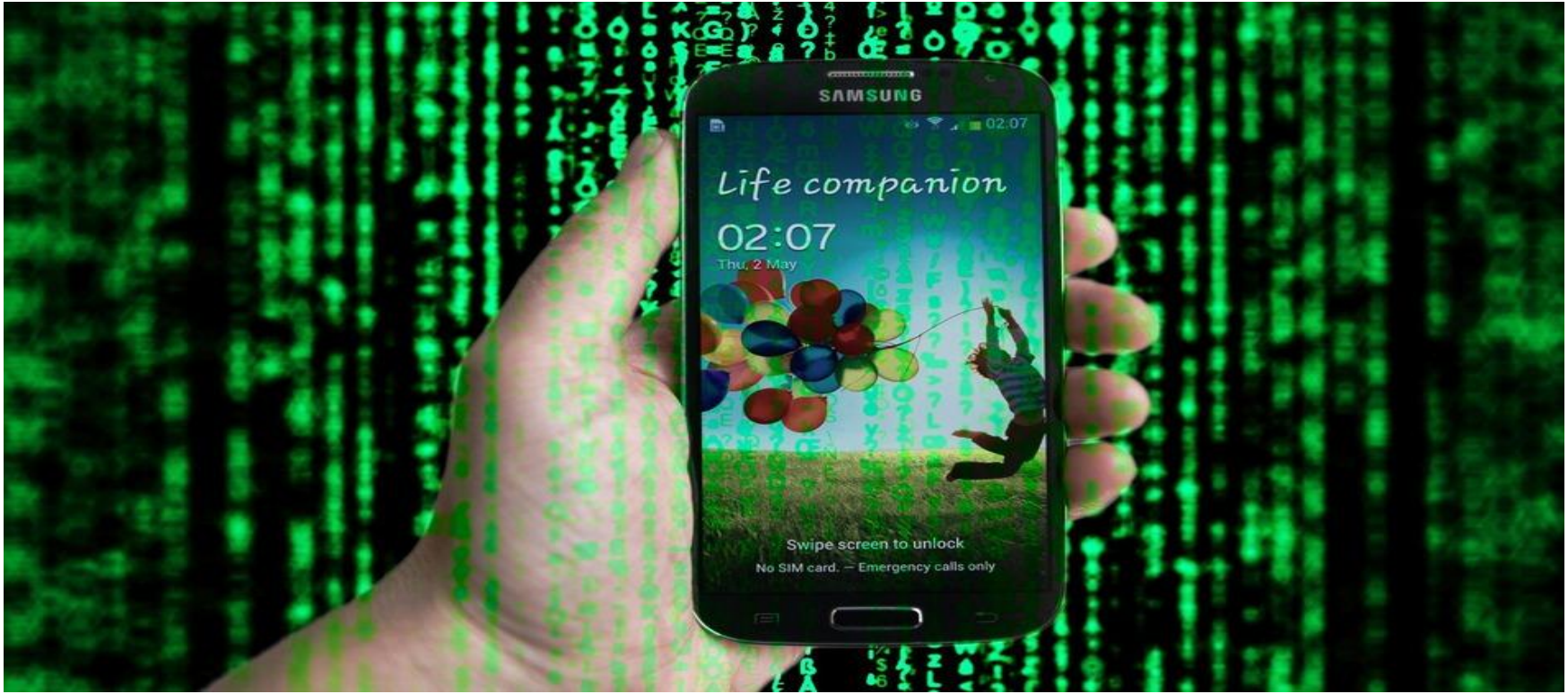
Tenth Anniversary Mobile Malware

- **2010**
  - Cross Platform Mobile Malware(Symbian and Android)
    - ZeusMitmo
- **2011**
  - Android
    - Geinimi, RootCager

Why are mobile devices attractive?



Life Companion



Dependence Mobile Devices

- **Business Purposes**
- **Everyday Needs**

Something for everyone!!

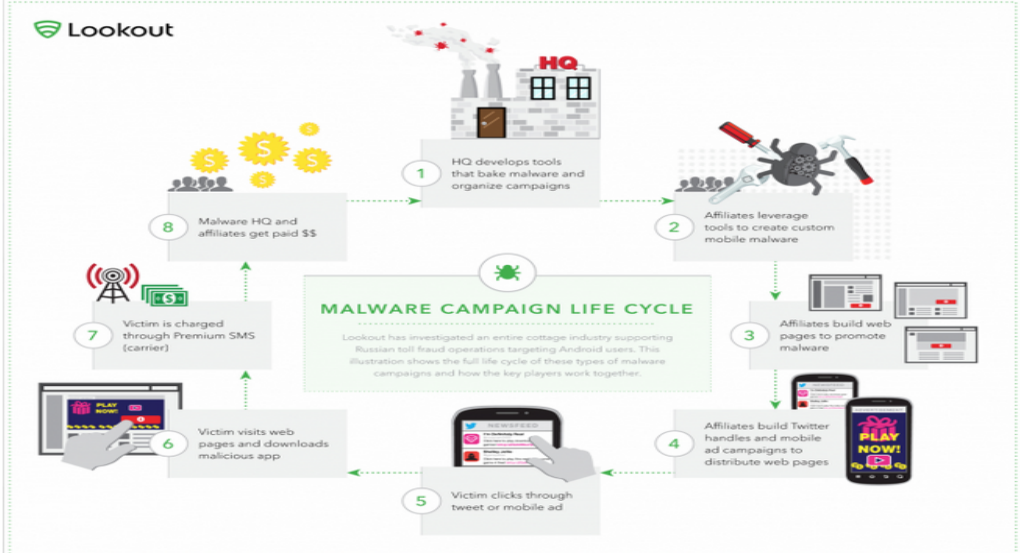
- **Cybercriminals**
  - Data Theft
  - Botnet Activity
  - Personal/Financial Information Theft
- **Government Entities**
  - Surveillance
  - Tactical Operations

Most Malware is not ‘new’

- **Repacks**
- **Incremental** Updates
- **Business Model**
  - Malware Headquarters
    - Startup Business
  - Governmental Intrusion and Remote Monitoring Solutions
    - Gamma International--FinFisher Suite

Malware HQ (Industrial Business)

- Dragon Lady Investigation
  - Lookout

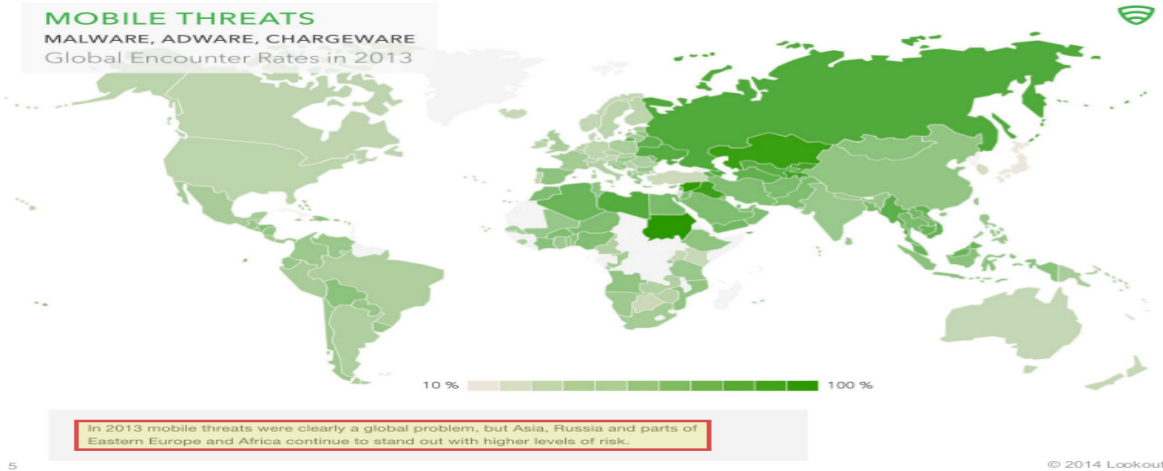


Dragon Lady Findings

- Android **Malware HQ**
  - Startup
  - Organized
- **Constant Releases** of Malware Families
  - Agile Approach
- **Affiliate Marketers**
  - Distribution
  - Customization

Why is this important?

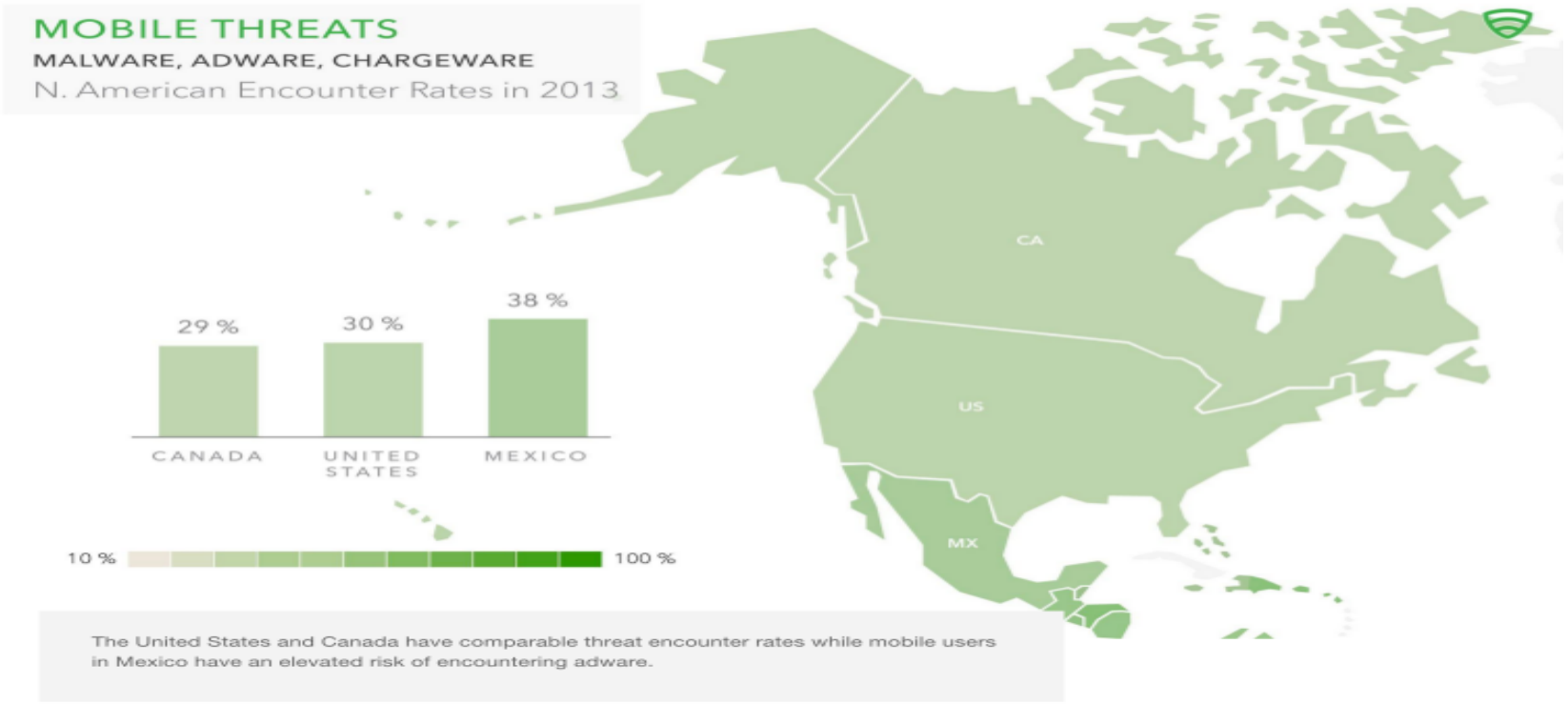
- Mobile **Malware Visibility**
  - Specific Geographical Regions (China and Russia)
    - Third Party App Stores



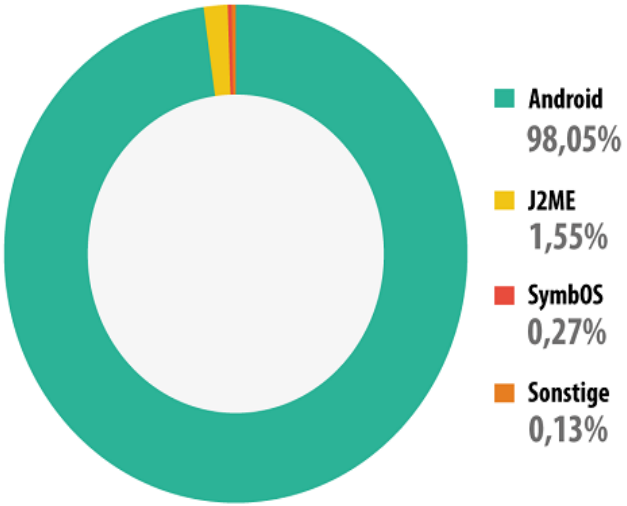
New Internet Citizens

- Experience the Internet through a **mobile phone**
- **Third World** Countries

North America

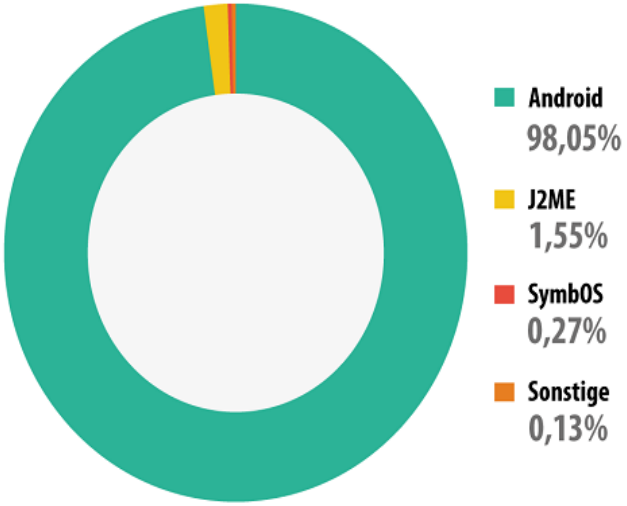


Malware Prefered Mobile Platform



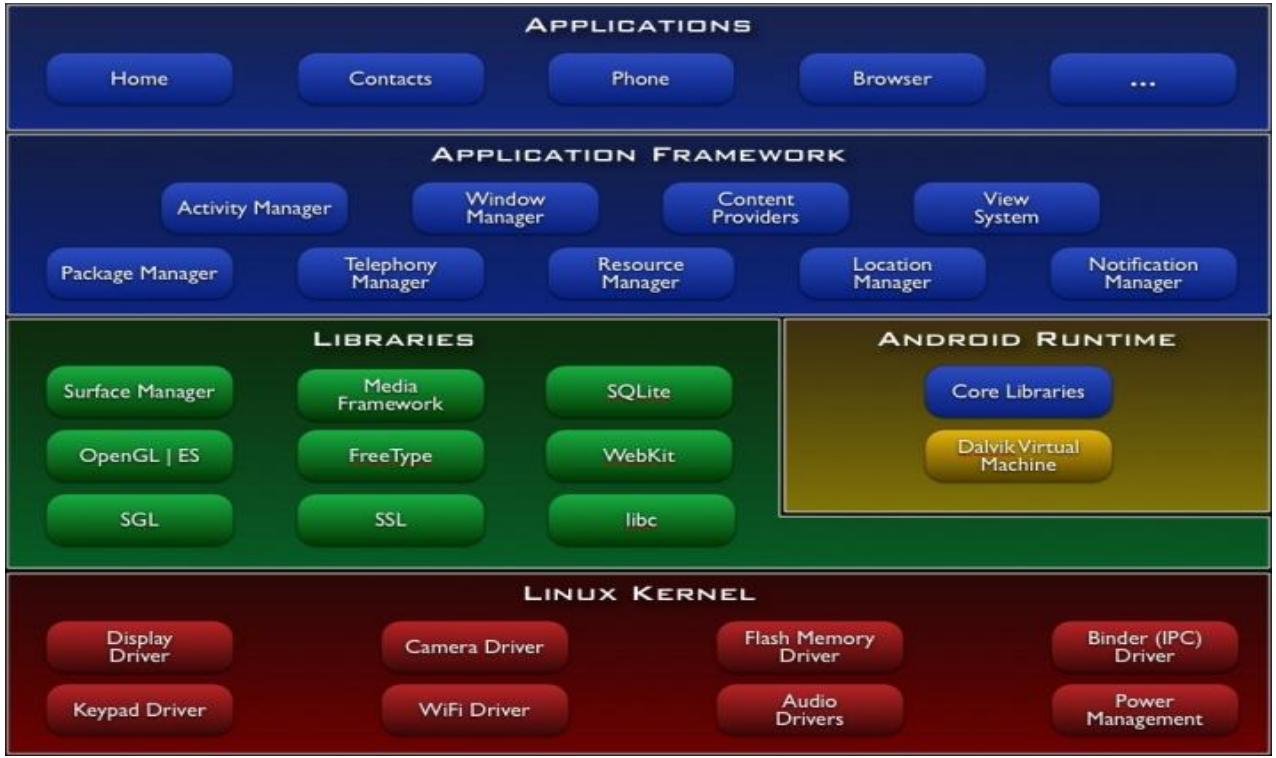
Mobile malware distribution by platform

Malware Prefered Mobile Platform



Mobile malware distribution by platform

Android Architecture



APK Package


.apk
Apps
Widgets
Calculator
Contacts
Updates
Clock
Email
Dropbox
Gmail
Gallery
Game Hub

Android Applications (APK)

- Zip Format Archive
  - **AndroidManifest.xml**
  - **classes.dex**
    - Java Code
    - Dalvik VM
  - **Meta information**
    - SSL Certificate (Self Signed)
  - **Resources**

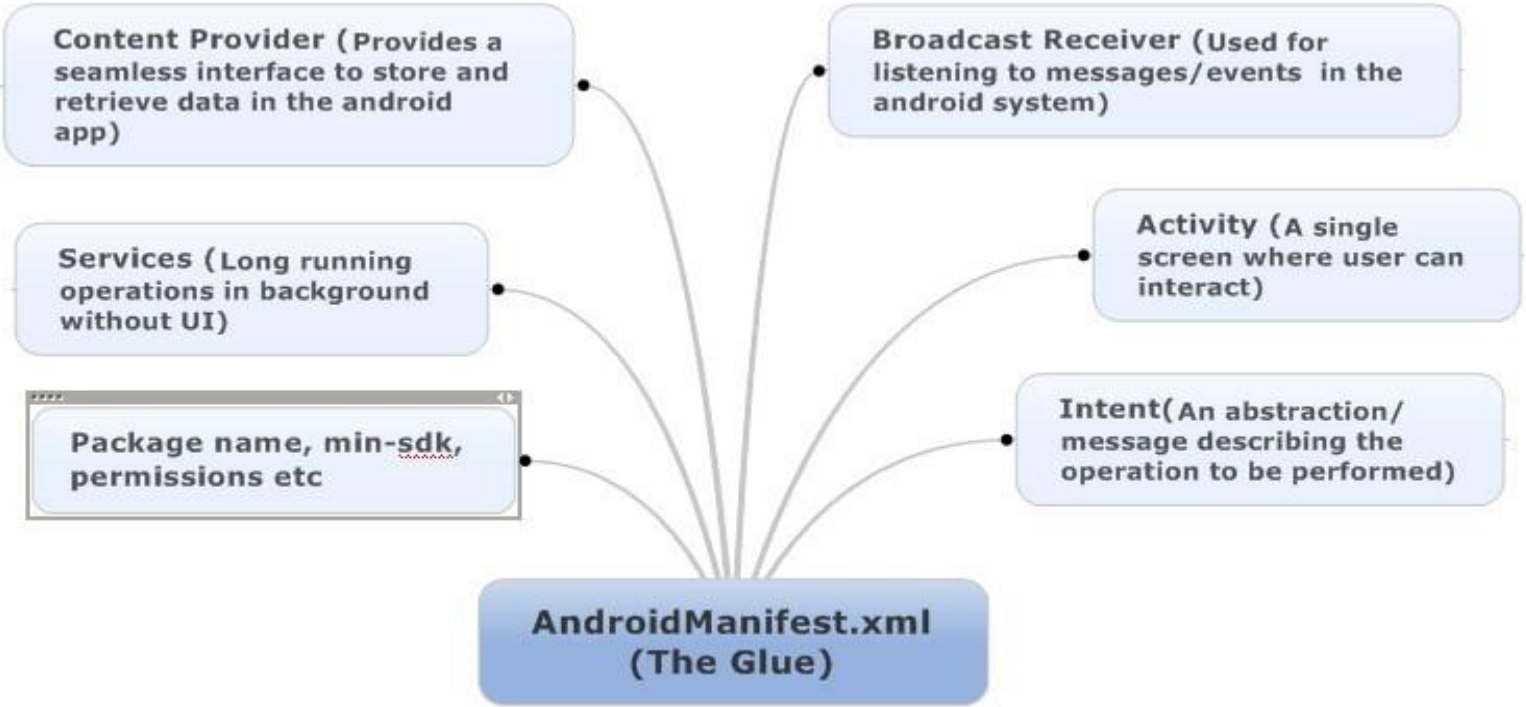
APK Internals

```
Archive:  Tor_58FED8B5B549BE7ECBFBC6C63B84A728_video.mp4.apk
  Length      Date    Time    Name
---------  ---------- -----   ----
     1651  2014-02-05 16:37   res/drawable/ic_launcher.png
   174398  2014-02-05 16:37   res/raw/debiancacerts.bks
   966562  2012-10-26 08:10   res/raw/geoip.mp3
   198652  2014-02-05 16:37   res/raw/iptables
   130204  2014-02-05 16:37   res/raw/iptables_g1
   134308  2014-02-05 16:37   res/raw/iptables_n1
   963309  2014-02-05 16:37   res/raw/obfsproxy
   346549  2014-02-05 16:37   res/raw/privoxy
      883  2014-02-05 16:37   res/raw/privoxy_config
  2074073  2012-10-26 08:10   res/raw/tor.mp3
      324  2014-02-05 16:37   res/raw/torrc
      291  2014-02-05 16:37   res/raw/torrctether
      480  2014-02-05 16:37   res/xml/policies.xml
     6104  2014-02-05 16:37   AndroidManifest.xml
     5648  2014-02-05 16:37   resources.arsc
     1651  2014-02-05 16:37   res/drawable-hdpi/ic_launcher.png
      887  2014-02-05 16:37   res/drawable-ldpi/ic_launcher.png
     1133  2014-02-05 16:37   res/drawable-mdpi/ic_launcher.png
     2597  2014-02-05 16:37   res/drawable-xhdpi/ic_launcher.png
  2917032  2014-02-05 16:37   classes.dex
    35890  2014-02-05 16:37   info/guardianproject/onionkit/trust/StrongTrustManager.java.underreview.txt
     4224  2014-02-05 16:37   ch/boye/httpclientandroidlib/impl/conn/tsccm/doc-files/tsccm-structure.png
     1795  2014-02-05 16:37   META-INF/MANIFEST.MF
     1848  2014-02-05 16:37   META-INF/CERT.SF
     1203  2014-02-05 16:37   META-INF/CERT.RSA
---------                     -------
  7971696                     25 files
```

Configuration Setup

Executable Code

AndroidManifest.xml

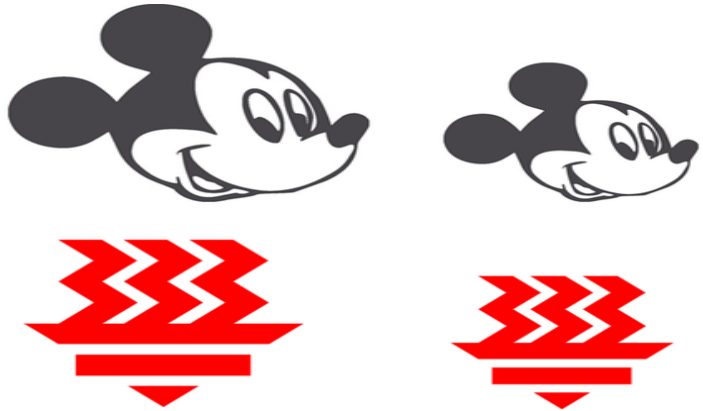


MEDS (Malware Evolution)

- **Malware Evolution**
  - **Android** Malware
  - Similarity Percentage
    - **Approximate Matching**
  - Creation Approximation
  - Phylogenetic/Lineage Tree

Similarity of object?

- Very good at **equality**
  - **Hashing (Fingerprint)**
- **Similarity of two objects?**



Approximate Matching

- NIST Special Publication **800-168**
- **Approximate Matching**
  - **Bytewise** (Sequence of bytes)
    - SDHASH
  - **Syntactic** (Internal Structures)
    - AndroidManifest.xml
  - **Semantic** (Contextual Attributes)

Phylogenetic (Lineage) Tree

- **Metadata**
  - Creation Date
- **Approximate Matching Value**
  - **AndroidManifest.xml (Syntactic)**
  - **Dex Files (Bytewise)**

MEDS (Discovery System)

- **Regression Analysis**
  - Feature Extraction
    - Number of Dangerous Permissions
  - Linear and Logistic Regression
    - Generativeness
      - Statistics about what malware will influenced the creation of future malware

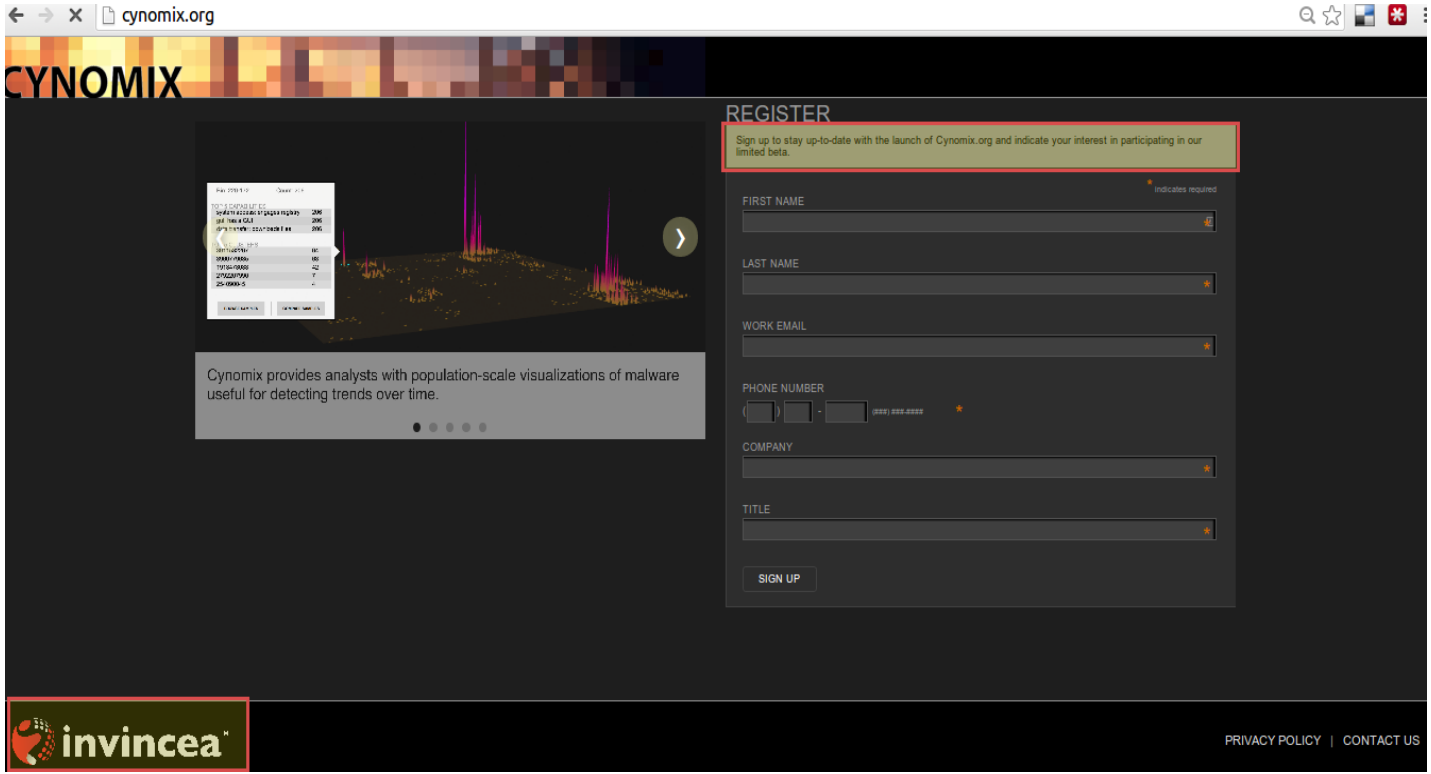
Python Implementation

- **Graphs**
  - **Pygraphviz**
    - Modeling
    - Dot Files for visualization
- **SDHash (Bytewise)**
  - Python SWIG Binding to C++ library
- **Edit Distance (Syntactic)**
  - AndroidManifest.xml

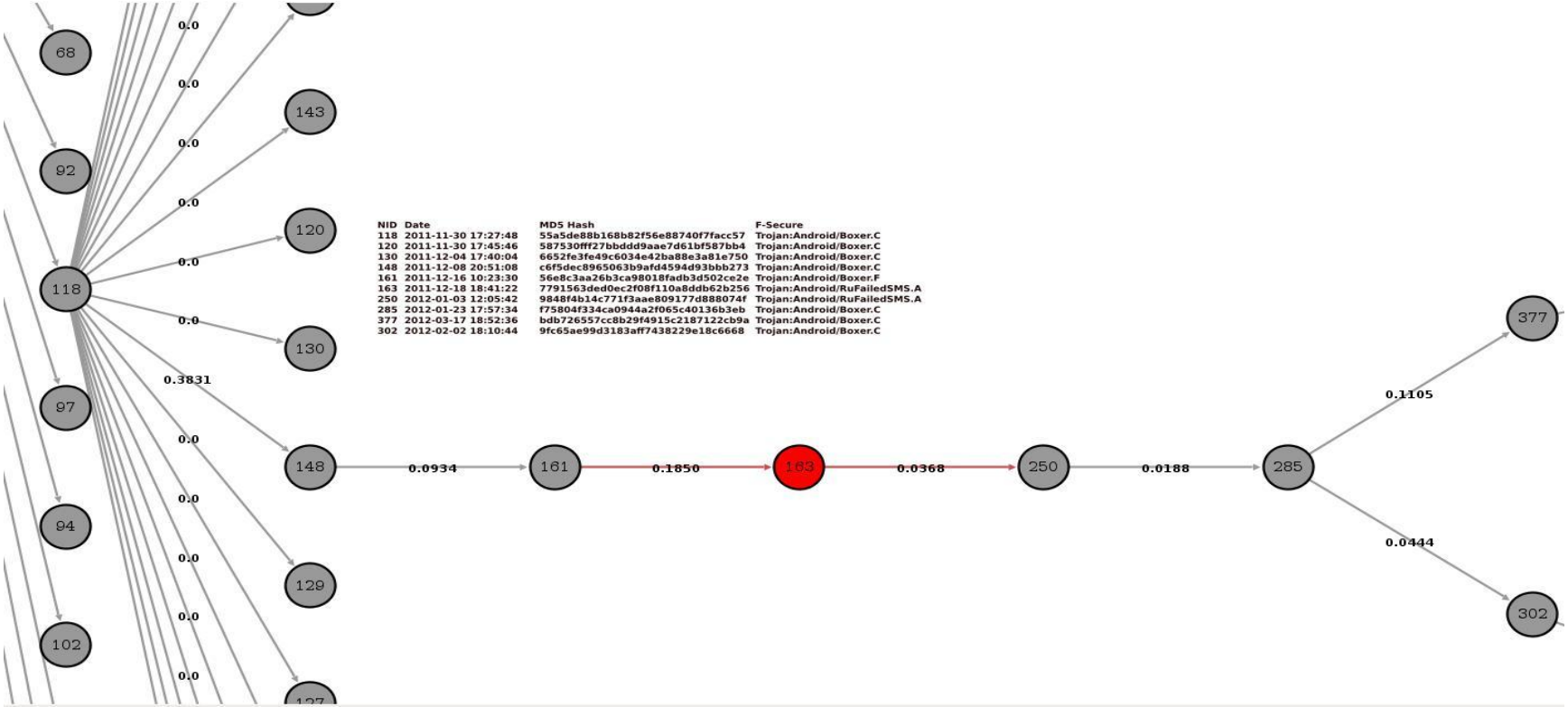
Phylogenetic (Lineage) Tree

- Not A New Idea
  - Goldberg, Leslie Ann et al. ***Constructing Computer Virus Phylogenies***. 1996
  - DARPA (**43 Million**) **Cyber Genome Project**, 2010
    - **Lockheed Martin**
    - **Invicea Labs** (Cynomix.org)
    - **BAE Systems**
    - **Raytheon BBN** Technologies

Invincea Labs (Cynomix)



Phylogenetic Tree



Generative Malware

- Generativeness
  - **Predict Future Malware Trends**
  - **Active Malware**
    - **Features**
    - **Vulnerabilities**
    - **Baseline**

Present Samples Scenario

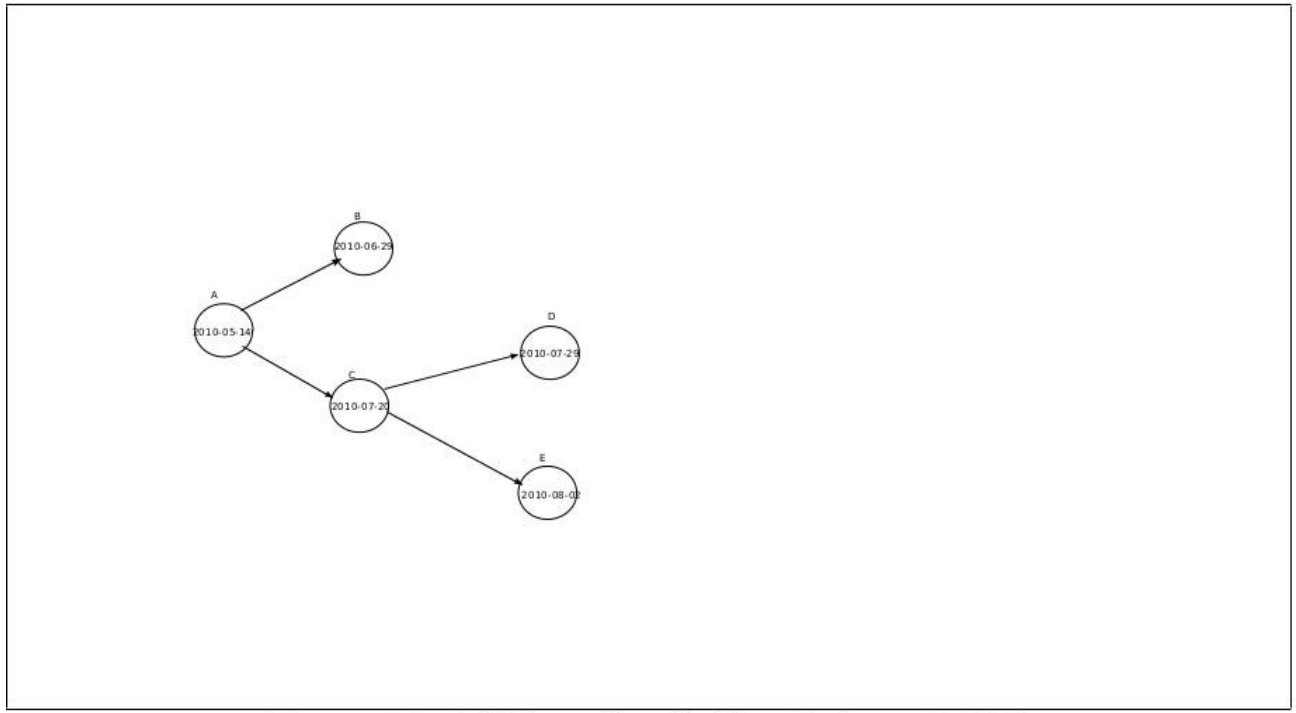


FIGURE 1.1: 5 Malware Samples May 14-August 2, 2010

Malware Evolution

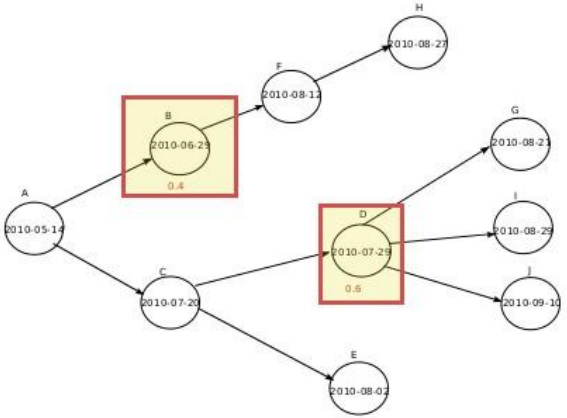


FIGURE 1.2: 10 Malware Samples Before and After August 2, 2010

Regression Analysis

- **Malware** Features
  - Number of Dangerous Permissions
  - Number of Receivers
- **Phylogenetic Tree** Features
  - Approximate Matching Value (to parent)
  - Age in second from parent
  - Age in second of the latest child

Dangerous Permissions

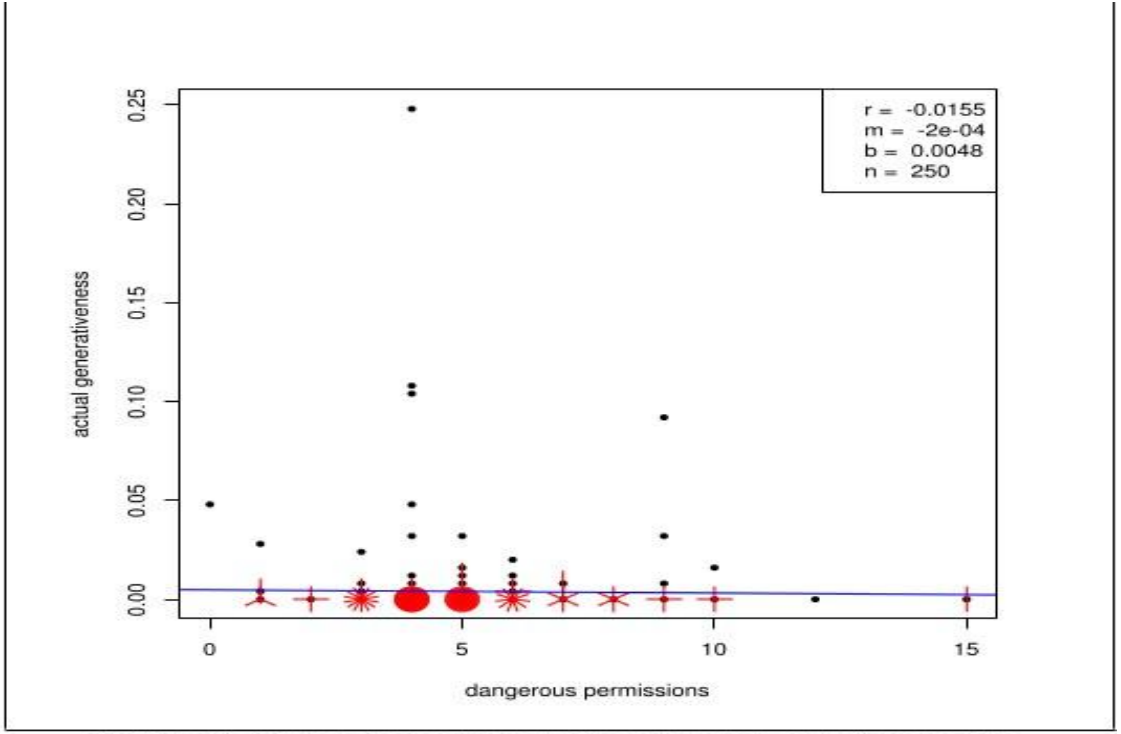


FIGURE 4.2: Number of dangerous permissions and actual generativeness values.

Number of Children

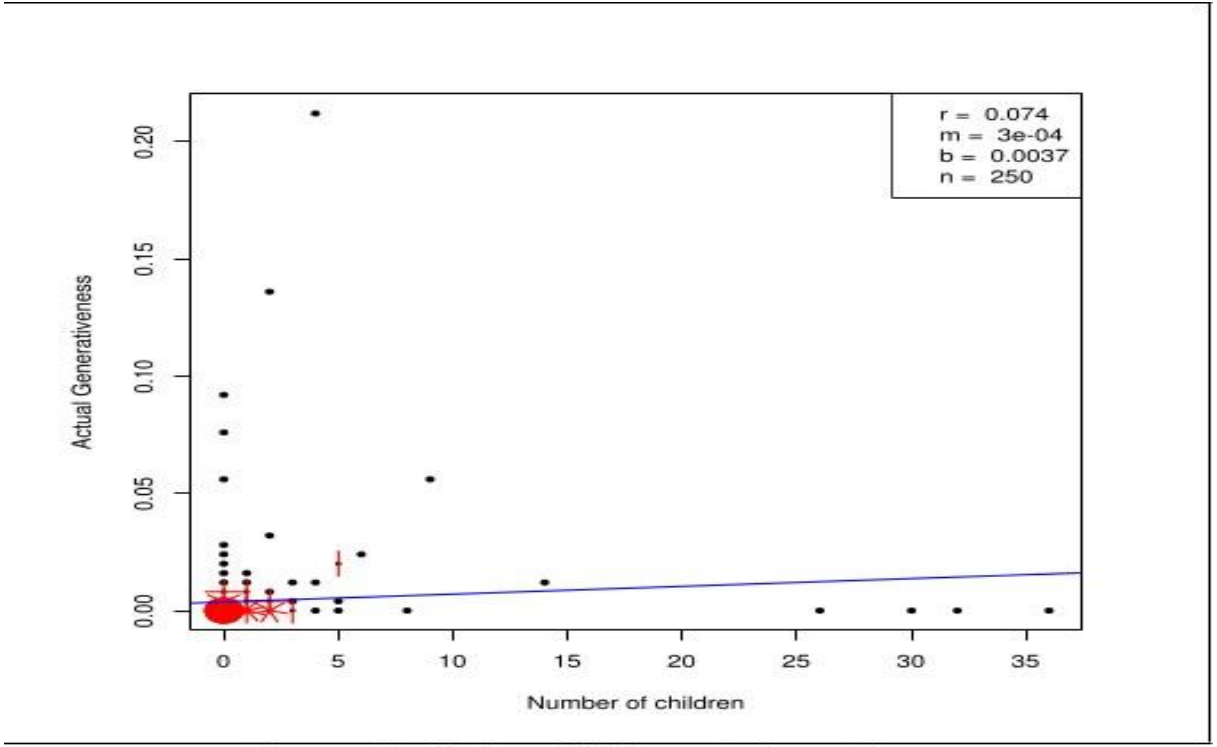


FIGURE 4.5: Number of Children actual generativeness

Python Implementation

- Regression Analysis Algorithms
  - **Octave**
    - Oct2py
  - **Visualization**
    - rpy2

Summary

- **MEDS**
  - **Phylogenetic** Tree Malware
    - Evolution Malware
      - Rapid Development of **detection** and **eradication**
  - **Generative** Malware
    - **Detect Promiscuous** Malware
    - **Pro-Active** Malware Outbreaks
    - **Data Science** Problem

Summary

- **Python very flexible**
  - Phylogenetic Malware Tree
  - Machine Learning Algorithms Integration
- **Generativeness (Data Science Problem)**
  - Bias
    - Further research
    - Choose Different Features

Future Work

- **Regression Analysis** improvements
  - scikit-learn
  - Bias Problem
    - Choose Different Features
    - Different Malware Sets
- Better **Visualization**
- More/Different **Malware Samples**

Acknowledgements

- **virusshare.com**
  - Malware Samples
- **Candice Quates**
  - SDHash Core Developer
- **CUNY**
  - **Prof. Bilal Khan**
    - http://www.systemic-inquiry.com
  - **Jeremy D. Seideman**

More Acknowledgments

- **AndroGuard**
  - Anthony Desnos
- **Silvio Cesare**
  - Software Similarity

Thank You !!!

- Questions
- Comments
- Clarifications

@vargasces